Lisbon and Porto Data Comparison

Maikel Pereira de Sousa

August 1st, 2019

1. Introduction

1.1 Background

Portugal is a country filled with different options for tourism and investments, every person that visits it wants to come back and even stay. For most, it's difficult to decide what to visit, depending on the season, some people might prefer some Portugal places than others. The two cities that always get thrown at this dilemma are Lisbon and Porto. Both cities have their particular beauty but for a small business investor or a tourist might be complicated to choose where to go.

1.2 Problem

Since the choice of the city of Lisbon or Porto might be a decision that takes multiple layers of study, we might want to see what kind of venues we would find as a visitor or investor. See where we'll have more competition or chances of success. This project aims to solve the question about the type of venues that are found in the different neighborhoods of the two cities.

1.3 Interest

Small and Medium size business might find this information useful to know where are the neighborhoods with less competition or where it makes sense to open a business, also, for tourism that might be looking to visit nonmainstream places of the two cities and see which neighborhoods can offer that.

2. Data acquisition and cleaning

2.1 Data sources

The base data that I use is on Wikipedia site, all containing data of postcodes of all the neighborhoods in Portugal. After that data is correctly extracted from the source gets completed with geolocation data from Google and Foursquare venues information.

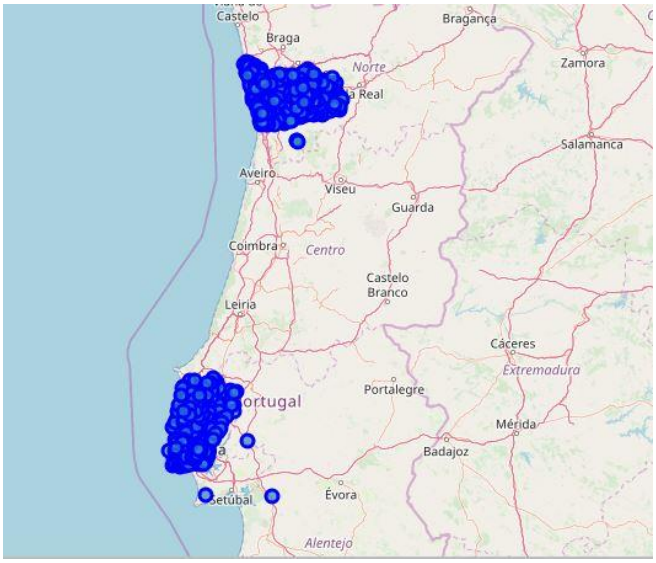
2.2 Data Cleaning

Data from Wikipedia is extracted onto an excel file since the website source code does not work well with packages like *beautifulsoup* for web scrapping.

After extracting that data in the excel file and creating a dataframe with the information about State, Postcode, Borough, and Neighborhoods in Portugal, I partitioned the dataset onto two dataframes with the information about Lisbon and Porto.

I complete the two dataframes with the latitude and longitude coordinates for each of the neighborhoods. Since the Google feature for some cases does not return a proper value I use different ways to retrieve it and even a deep feature to know how many times the feature was called.

To be sure that the information is extracted correctly I graph the latitude and longitude points on a map and as a result, some minor data must be dropped.



Before dropping

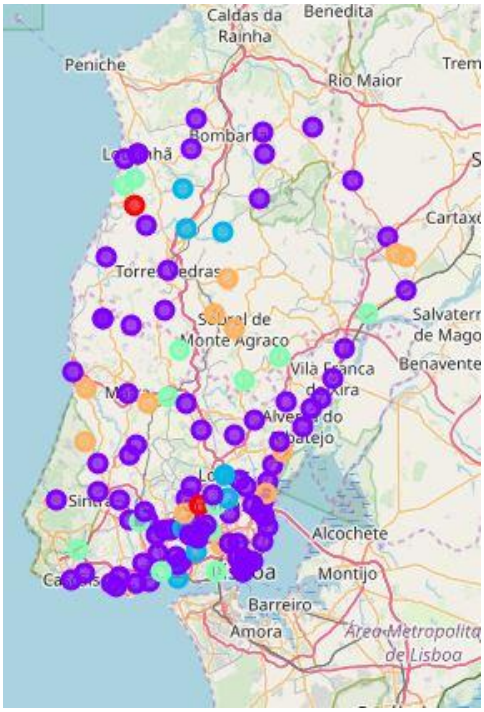
After dropping

2.3 Feature selection

With these two datasets of size (120, 11) for Lisbon and (105, 11) for Porto, I can start our search, in Foursquare API, of venues in a 700 meters radius. That information gets classified by the type of venues and then I study the top 10 venues in each neighborhood.

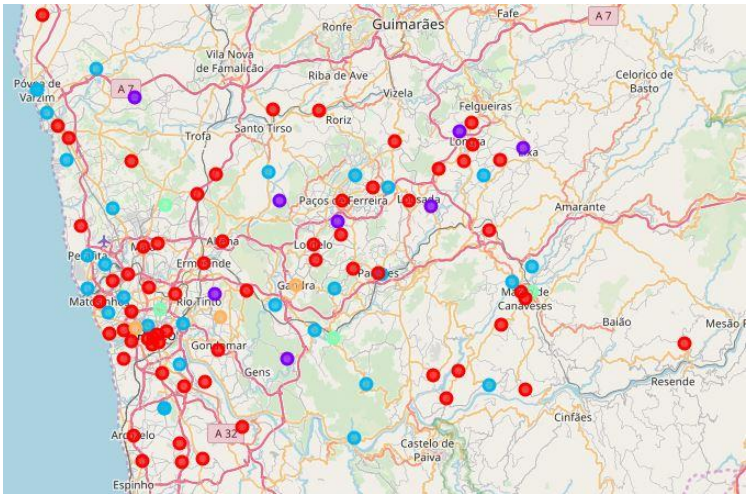
Although some in some neighborhoods I get no venue information, those places get discarded from the analysis. By looking this into a map I have:

For Lisbon:



Each color represents the cluster that the neighborhood belongs to

And, for Porto:



Each color represents the cluster that the neighborhood belongs to

3. Exploratory Data Analysis

3.1 Clustering

At the plain sight, the information about the top 10 venues in each city is not clear enough to make assumptions. A clustering algorithm is needed, a K-means algorithm with k=5 is the choice for each city.

3.2 K-means classification

The data for each cluster gets properly distributed and the type of venues in each dataframe is visually distinguishable from others and between cities.

Some are more related to the local type of venues:

	Postcode	deep	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
94	4475	1	3.0	Breakfast Spot	Grocery Store	Diner	Electronics Store	Farmers Market	Fast Food Restaurant	Flea Market	Flower Shop	Food	Garden
108	4630	1	3.0	Breakfast Spot	Supermarket	Wine Shop	Diner	Electronics Store	Farmers Market	Fast Food Restaurant	Flea Market	Flower Shop	Food
177	4585	1	3.0	Breakfast Spot	Grocery Store	Diner	Electronics Store	Farmers Market	Fast Food Restaurant	Flea Market	Flower Shop	Food	Garden
205	4200	1	3.0	Breakfast Spot	Grocery Store	Diner	Electronics Store	Farmers Market	Fast Food Restaurant	Flea Market	Flower Shop	Food	Garden

Or completely tourist neighborhoods:

	Postcode	deep	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	2630	1	3.0	Portuguese Restaurant	Café	Wine Bar	Electronics Store	Frozen Yogurt Shop	French Restaurant	Food	Flower Shop	Fast Food Restaurant	Farmers Market
1	2630	1	3.0	Portuguese Restaurant	Pharmacy	Restaurant	Café	Dumpling Restaurant	Food	Flower Shop	Fast Food Restaurant	Farmers Market	Electronics Store
10	2050	1	3.0	Portuguese Restaurant	Wine Bar	Electronics Store	Frozen Yogurt Shop	French Restaurant	Food	Flower Shop	Fast Food Restaurant	Farmers Market	Dumpling Restaurant
23	2755	1	3.0	Portuguese Restaurant	Seafood Restaurant	Snack Place	Electronics Store	Frozen Yogurt Shop	French Restaurant	Food	Flower Shop	Fast Food Restaurant	Farmers Market
29	1300	1	3.0	Portuguese Restaurant	Basketball Stadium	Wine Bar	Farmers Market	Frozen Yogurt Shop	French Restaurant	Food	Flower Shop	Fast Food Restaurant	Electronics Store

4. Results

Our analysis shows that, although there are a lot of venues for each of the two cities, Lisbon is the one with more venues even when Porto has more neighborhoods. I know this, by comparing the sizes of the dataframes before and after the foursquare search. Lisbon reduces 29% approx. but Porto 67% approx.

The reason for this is that those places without geolocation data are mainly residential areas with no venues or little venues irrelevant for the users, therefore making those irrelevant to foursquare. As a result, I can see that Lisbon has

neighborhoods intended for different purposes but Porto shares more uniformity in its neighborhoods.

Our k-means algorithm partitions our universe of venues in the best way. Every postcode has similar venues and, in some, identical distribution.

By looking at the dataframes in both cities the type of venues are mainly related to food and the same clusters expand from the center, although for Porto that uniformity extends beyond the city center.

5. Conclusion

The purpose of this project was to find similarities and differences between Porto and Lisbon. Each city has its particularities.

For the city of Porto, the venues are less in number when compared to Lisbon but show uniformity on all its extension, and seems a city intended for visitors that are looking for a complete Portuguese experience. In the case of Lisbon, we find a city that is prepared to deliver different experiences to its visitors or investors. Uniformity of venues is something that is also found but in certain particular cases close to downtown but it fades and changes as you move around it.

Both cities show their particular potential, it depends on the particular investor to decide in which city its investment has more impact, for food investments that aren't related to Portugal the city of Porto shows lots of potentials since there's no competition in the area but if you are planning to deliver services to Portuguese people Porto is going to be your go-to place since it is a place with lots of locals, but consider that you might lose in the variety of business opportunities.

On the other hand, Lisbon is a city where you'll find more competition in certain areas, including local and foreign food venues but not enough to saturate the market. By doing a more detail study good spots for food investments can be found but Lisbon shows a lot of potential in another type of venues outside the food sector.

To close this project, it is fair to say that these two cities show lots of potential in the development of food sector innovations, whether be for locals or visitors. One could be more open to new venue proposals but in Porto is important to be cautious in the type of innovations, or at least, be aware that is a city with some uniformity in the venue types.